AFRL-HE-BR-TP-2000-0006

UNITED STATES AIR FORCE RESEARCH LABORATORY

STRENGTH AND RESISTANCE TO INTERFERENCE IN PRACTICED RECOGNITION: MEMORY-RETRIEVAL ABILITIES INVESTIGATED THROUGH LATENT STRUCTURE MODELING

Scott R. Chaiken

AIR FORCE RESEARCH LABORATORY
HUMAN EFFECTIVENESS DIRECTORATE
WARFIGHTER TRAINING RESEARCH DIVISION
BROOKS AFB, TEXAS 78235-5118

July 2000

NOTICES

This report is published in the interest of scientific and technical information exchange and does not constitute approval or disapproval of its ideas or findings.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Scott R Chaiken

SCOTT R. CHAIKEN
Project Scientist

William E Wimpee
WILLIAM E. WIMPEE, Lt Col, USAF
Chief
Information Systems Training Branch

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE
July 2000

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Strength and Resistance to Interference in Practiced Recognition: Memory-Retrieval Abilities Investigated through Latent Structure Modeling

5. FUNDING NUMBERS
PE - 61102F
PR - 2313
TA - HM
WU - 15

6. AUTHOR(S)
Scott R. Chaiken

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
Brooks Air Force Base, TX 78235-5118

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-HE-BR-TP-2000-0006

11. SUPPLEMENTARY NOTES

Air Force Research Laboratory Technical Monitor: Dr. Scott Chaiken, 210-536-2870

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200* words)
In support of improving general research practices and USAF selection procedures, AFRL/HEAI has documented AFOSR-sponsored research leading to the acceptance for publication of "Memory-retrieval abilities investigated through latent-structure modeling" in the American Journal of Psychology (expected 2001, vol. 114). While the paper examines issues in cognitive and differential psychology (a necessity for publishing in this kind of journal), the AF relevant part of the article is methodological. The paper shows how a criterion can be decomposed, by careful design of part tasks, into a set of hypothetical abilities important for the criterion. Latent structure modeling is then applied to assess both the reality and the importance of the underlying abilities to the criterion. This methodology provides a structured, potentially job-tailored, approach to selection test construction and validation. This work is legacy work from closing work units. There is no active selection research at AFRL.

14. SUBJECT TERMS
Memory, Latent-structure models, Individual differences, Cognitive psychology

15. NUMBER OF PAGES
39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 3-00)
Prescribed by ANSI Std Z-39-18
298-103 COMPUTER GENERATED

Table of Contents

Abstract

Participants ($\underline{N}$=811) practiced paired-associate recognition with and without an interference manipulation and then practiced a pattern-recognition skill where patterns discriminated had features in common. Structure models of the covariances among task reaction times indicated two factors or abilities. The first was a baseline factor, hypothesized to include the ability to strengthen traces and other abilities common to all tasks. The second was a resistance-to-interference factor, or the ability to quickly retrieve associations with elements in common with non-retrieved associations. Further modeling on a subset of the sample ($\underline{n}$=434) showed the baseline factor to reflect a memory-strength ability, independent of other confounding abilities (e.g. motor, reading abilities). Both memory abilities are discussed broadly with respect to cognitive skill acquisition, controlled vs. automatic processing, and activation.

Strength and Resistance to Interference in Practiced Recognition: Memory-Retrieval Abilities Investigated through Latent Structure Modeling.

Manipulations of memory strength typically involve variations of practice or exposure on materials composing a single memory trace. Such manipulations may reflect the quality of encoding operations. Manipulations of interference typically involve practice or exposure on materials composing multiple traces sharing elements. Such manipulations may reflect the efficiency of retrieval operations. These two constructs often show up as distinct parameters in general memory theories (e.g. image-strength vs. sampling probability in Gillund and Shiffrin's, 1984, SAM model, node strength and fan in Anderson's 1983, ACT* model). Given historically distinct manipulations and theoretical roles for strength and interference in the modeling of memory tasks, these constructs could reflect different mechanisms or stages of processing within the individual.

Evidence that memory strength and interference affect different stages of information processing can be found in, at least, two types of empirical study. One source is studies that find additive effects of strength and interference manipulations (Anderson, 1981; Gillund and Shiffrin, 1984; Howe, 1995). Such effects suggest different stages of processing are affected (Sternberg, 1969). Another source is speed-accuracy trade-off studies. When different parts of the speed-accuracy function are affected by manipulations of strength and interference, it implies these manipulations affect different stages (e.g. Dosher 1981, 1984).

One can also assess whether memory strength and interference reflect different processing stages by assessing whether these manipulations reflect <u>different</u> types of human ability. Presumably, if performance under different conditions depends on different (i.e. somewhat independent) abilities, performance under these conditions would employ different stages of processing as well. An "individual differences" demonstration of a stage is orthogonal to the other sources cited above. In other words, additive factors or speed-accuracy results do not imply anything definitive about the <u>individual differences</u> present in different stages. For instance, it is possible to have additivity of an effect with the absence of individual differences in the effect (i.e. a significant factor effect with the absence of subject-by-factor interactions). One could also have significant individual-differences in each (additive) factor, but have the individual-differences observed for factors correlate perfectly (as with a more global factor affecting multiple distinct stages, Jensen, 1987).

To observe distinctness in memory strength and interference as <u>abilities</u>, one must show that performance rankings differ among people, in a quantifiable sense, depending on what the task depends on (i.e. memory strength or interference). One must also show that these rank differences are <u>reliable</u> in some sense, or have some generality across different tasks thought to depend on the hypothetical ability. Such observations depend on psychometric or correlational observations, which do not depend on the mean data.

The recent emergence of latent-structure modeling routinely allows the observation of different types of ability in the sense just described. That is, one can demonstrate two classes of tasks are composed of two different abilities, by assessing their data in the context of models with and without the assumption of two abilities. However, an initial prerequisite is finding a set of tasks that differentially relate to the hypothesized abilities, memory strength and resistance-to-interference ability. I have chosen a pair-recognition paradigm, because that paradigm was amenable to constructing such tasks.

Experimental Strategy

The general hypothesis tested in Experiment 1 was that recognition memory tasks with an interference component could not be perfectly explained by other recognition tasks hypothesized to rely mainly on memory strength. This finding would be expected under the assumption that recognition tasks with interference involve some unique ability above and beyond the ability to accumulate memory strength. I used two pair-recognition tasks to experimentally deconfound a resistance-to-interference ability from an ability to passively accumulate memory strength. I used a third task, procedural learning, to demonstrate the generality of the strength-independent interference ability across two different types of task hypothesized to share (the distinct) interference ability.

Specifically, structure modeling was employed as a type of theoretical regression. Two underlying latent factors (i.e. abilities) were posited as the independent variables for predicting pair recognition and procedural learning reaction times. Task reaction times were then regressed on these two latent factors in order to fit the observed variance/covariance matrix for all reaction time scores. Of interest was whether models with two ability factors fit better than models with only one factor. A one-factor model would be implied if ones susceptibility to interference were totally determined by ones ability to accumulate memory strength.

This type of structure modeling is similar to multivariate regression in which observed variables are predicted by other observed variables (c.f. Long, 1983). One might wonder whether theoretical (latent) variables should be preferred to observed predictors in a multivariate regression analysis. The chief benefit for latent variables is protection against certain kinds of measurement ambiguity. In the typical case where all variables (predictors and the criterion) are observed, independent (or incremental) prediction for two variables predicting a third could mean that the two variables measure different (independent) components of the third. However, it could also mean that the two predictors measure the same component of the third but less reliably alone than in combination with each other (leading to significant regression betas for each predictor in the same standardized regression equation predicting the criterion).

However, latent factors do not have this measurement ambiguity. Because latent structure models estimate both how much an observed score depends on its uniqueness from other scores (i.e. its "error") and how much it depends on shared factors, reliability of the score is part of the model. Furthermore, the factors themselves, as modeled entities, have perfect reliability. Hence, when two factors independently predict an observed score, this unambiguously implies multiple sources of ability (or variance) for that score. Of course, one caveat to this modeling boon is that the latent-structure model has to be "correct" or "true" in some sense. In practice, this caveat is assessed to the extent competing models have been shown inferior.

Figure 1 (Model 1) shows a two-factor model for tasks employed in Experiment 1. Of principal interest are the arrows originating from the factors (ellipses) which terminate at observed scores (labeled boxes). The factor-to-score arrows represent the regression betas of latent factors on observed scores (e.g. performance on the 5th epoch of the procedural learning task, performance on the 4th alternate form of pair recognition under interference). These betas provide an index of the effect size for a factor (described later in the discussion to Experiment 1).

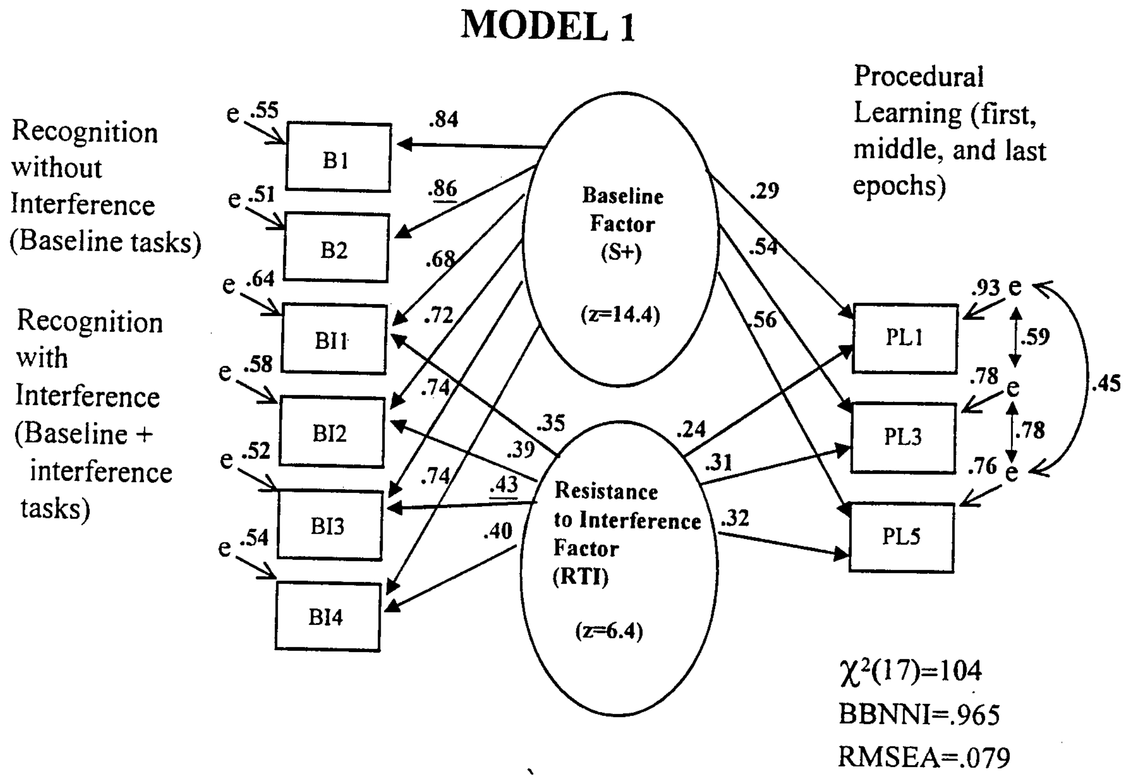

Figure 1.

Two factors in practiced recognition, a baseline factor including memory strength, and a factor for resisting interference. Score abbreviations are B1, B2, Baseline (no-interference) tasks; BI1 - BI4, Baseline+interference tasks; PL1, PL3, PL5 procedural learning epochs (early, middle, and late). All standardized betas (factor paths) are significant ($\underline{z} > 3.0$), except underlined path estimates whose unstandardized weights were fixed at 1.0 in order to free the factor variance parameters ($\underline{z}$s for which are also displayed). Fit statistics are Bentler-Bonett Nonnormed fit Index (BBNNI) and Root Mean Square Error of Approximation (RMSEA). $\underline{n}$=811.

The figure shows the underlying cause of the correlations between tasks to reside (in part) from separable interference and baseline abilities. Tasks presumed to have no interference only correlate to other conditions via a baseline factor common to all tasks. This baseline factor includes a memory-strength factor, but other factors as well (hence S+ in the figure). Other tasks are presumed to reflect two underlying factors, baseline and resistance to interference. The interference factor is "nested" (c.f. Gustafsson and Balke, 1993) in the baseline factor, in the sense that the interference ability is unique to a subset of tasks, which also depend on baseline processes. Nested factors may be thought of as factors existing on the "residual" of a score after prediction by the factors they are nested within. In other words, Model 1 assesses whether there is any reliable variance left over after the effects of the baseline factor have been accounted for.

If another factor than the baseline is required in Model 1, the additional factor can be attributed to an interference ability by virtue of the experimental design. Specifically, one of my tasks has been designed to differ from baseline conditions only by the addition of an interference manipulation. Hence, when the baseline-plus-interference task and some other task (i.e. the procedural learning task) are found to share variance above what can be accounted for by the baseline factor, the conclusion is that the new ability was "introduced" with the modification of the baseline task. This general tactic of defining an ability by an extension of a baseline task is a frequent stratagem in the individual-differences literature (e.g. Kyllonen and Tirre, 1991; Sternberg, 1977; Sternberg and Gastel, 1989; Tirre and Pena, 1993; Woltz, 1988; Yee, Hunt, and Pellegrino, 1989). There have also been more formal discussions of the value of this technique (e.g. Donaldson, 1983).

The baseline task has no interference effects associated to it within the model. I argue that my implementation of the baseline task is what actually achieves this, but alternatively, I could say that the model I will explore will assess the reasonableness of this conjecture for the baseline task. The baseline task was a word-pair recognition task in which pair words occurred only in one response category. Hence, participants only accumulated memory strength on word-to-response associations as they practiced recognizing pairs. In addition to memory strength (ability), this task plausibly contains speed of basic motor response and letter/word reading (abilities). The baseline+interference task was identical to the baseline task except for the attribute of interference. Specifically, the baseline+interference task had each word of a pair occur in both response categories. Hence, participants used the particular combination of words to determine the response, and interference arose from multiple use of the same words in different pairs requiring different responses.[1]

One could demonstrate the existence of interference ability in recognition reaction time (hereafter, RT) just by modeling correlations among baseline and baseline+interference pair-recognition tasks. However, my intent was also to examine the generality of the interference construct beyond simple pair recognition. Therefore, the procedural learning task from Woltz, 1988, was also considered. Unlike the pair-recognition tasks, which used different materials with each replication, the procedural learning task is a rule-based categorization task on a large set of repeated materials. Substantial learning has been shown in this task (Woltz, 1988). Hence, within the latent-structure modeling one can also look at the effects of strength and interference ability in the context of acquiring a simple cognitive skill.

The learning context is a side issue from the primary one of determining whether strength and interference operations reflect two different abilities. However it is an important side issue that helps to clarify the nature of the abilities studied. If resistance-to-interference shows independence from strength (i.e. two factors are necessary), then that independence could reflect either a controlled or an automatic processing ability (or possibly a combination of both). I won't rigorously define automatic vs. controlled processing other than referring to an ecologically valid diagnostic related to skill acquisition. If resistance-to-interference is primarily a controlled process, it should decline in its importance to procedural learning as the latter is practiced (Ackerman, 1988; Woltz, 1988). However, if resistance-to-interference is primarily an automatic processing ability, its importance to procedural learning should increase as the latter is practiced. There are reasons for expecting a greater impact of a resistance-to-interference ability (should such exist) both in the early and late phases of the procedural learning, as I develop next. The next section will also explain why I chose the procedural learning task as a different type of memory task (from pair recognition) that is limited by both memory strength and interference.

Cognitive Skill Acquisition

In Woltz (1988) "procedural learning" was used to investigate how information-processing abilities changed in a cognitive skill as it was practiced. Hence, Woltz designed procedural learning to be like a typical cognitive skill, albeit with simple rules so the task could be rapidly learned in the laboratory. Consistent with this characterization, participants progressed from slow and error-prone problem solving to fast and accurate performance with modest practice. Decreases in RT also followed the power-law function typical in skill acquisition (Newell and Rosenbloom, 1981).

In procedural learning participants classify numbers in varying format by applying pre-memorized rules:

"If a number is in WORD-form, check whether the number is ODD or EVEN. If ODD and in the LOWER half of the screen or EVEN and in the UPPER half, press the RIGHT button. For any other WORD configuration press the LEFT button.

If a number is in DIGIT-form, check whether the number is BIG (>10) or SMALL (<10). If the number is SMALL and in the UPPER half of the screen or BIG and in the LOWER half press the RIGHT button. For any other DIGIT configuration press the LEFT button."

Early in skill learning, reading out and interpreting these complex propositions was hypothesized to tax working memory, but as learning progressed working memory would pose less of a limitation. Woltz (1988) demonstrated this by showing correlations of working memory to procedural learning declined with practice on the latter. This was hypothesized to be a consequence of the skill becoming production-based with practice. Some reasonable productions for practiced procedural learning are (after Woltz, 1988):

If <u>WORD, ODD, and LOWER HALF</u> are present, press the RIGHT BUTTON.

If <u>WORD, EVEN, and LOWER HALF</u> are present, press the LEFT BUTTON.

If <u>WORD, ODD, and UPPER HALF</u> are present, press the LEFT BUTTON.

If <u>WORD, EVEN, and UPPER HALF</u> are present, press the RIGHT BUTTON.

Woltz (1988) also hypothesized that production-based performance should depend critically on the ability to strengthen the associations between condition elements of productions (italicized above). He demonstrated this by showing that a person's memory-strengthening ability related to late procedural performance better than early procedural performance. This was in contrast to working memory, which related most clearly to early performance. Memory-strengthening ability (then called an "activation savings" ability, c.f. Woltz and Shute, 1993) was measured by the amount of repetition priming exhibited in another task.

The current study advocates resistance to interference as another memory-retrieval ability relevant to late skill performance. Interference can limit speed late in a skill given such effects can be found after extended practice (Pirolli and Anderson, 1985) and given the form productions for the procedural learning task are expected to take with task practice. Such productions should share elements that have conflicting responses associated with them (e.g. the chunk word/odd/upper-half which matches a production for pressing left and the chunk word/odd/lower-half which matches a production for pressing right). Anderson's 1983, 1993 memory-retrieval models would predict this sharing of elements to slow pattern-matching for reasons similar to that described for fan-effect paradigms (Anderson, 1983). In fan-effect paradigms, recognition RT for memorized sentences is longer for sentences composed of concepts used in other learned sentences. The frequency of use of a concept in distinct sentences is what corresponds to the concept's fan.

However, resistance-to-interference could also have a large effect early in procedural learning, given Woltz's (1988) finding that working memory is important to early performance, and given recent findings suggesting interference in recognition and working memory are sometimes related. Conway and Engle (1994) have shown some fan effects to be sensitive to working-memory capacity while other fan effects are not. The baseline+interference task, considered as a type of fan task, would seem to belong to their working-memory class (as I argue later). Hence, one might expect any distinct interference ability demonstrated beyond strength and other baseline abilities to reflect a type of "controlled processing" or limitation of "attentional" resources to use Conway and Engle's (1994) terms.

Experiment 1: Two Factors in Practiced Recognition

Method

Summary of data collection studies

Data collection for Experiment 1 occurred in three studies—A (n=179), B (n=193), and C (n=434). Data collection for Experiment 2 was from Study C. With respect to measuring a resistance-to-interference ability independently from strength and other baseline abilities (Experiment 1), the 3 studies can be aggregated as they used identical procedural learning and pair-recognition tasks. Study A and B differ only in the ordering of pair-recognition tasks and in the inclusion of some unique filler tasks coming between pair-recognition tasks. For Study A, baseline tasks were given before baseline+interference tasks. For Study B the reverse was true. For Study C ordering was balanced. Study C is the best design from the perspective of separating a strength ability from other confounding abilities in the baseline tasks (Experiment 2). That is in addition to the baseline and baseline+interference tasks, Study C contains tests relevant to measuring these confounding abilities (while Study A and B do not). There was also a Study D (n=478), not included in Experiment 1 and 2 analyses. This study was a close replication of the procedure and results of Study C (Experiment 2). This sample is not included because Armed Services Vocational Aptitude (ASVAB) scores were unavailable (due to the timelag in getting such scores and not because the sample was special). Such scores were critical to an analysis in Experiment 1. However, I provide more on Study D below (footnote 5). Study D is of interest as it empirically shows the replicability of a fairly complicated latent-structure model applied to a new sample.

Participants

Participants were 934 (81% male) Air Force basic trainees. However, owing to missing data for reasons described in the results, 811 participants were analyzed. All participants scored at or above the 40th percentile on the Armed Forces Qualifying Test, all had finished high school (or the equivalent) and had vision corrected to Air Force standards.

Apparatus

All experimental tasks were given on 486 50 MHz or higher computers housed in library-style carrel in a single-room with 40 testing stations. All computers used mouse, keyboard, and 15-17 inch color monitors, with an approximate viewing distance of about 65 cm. Reaction time measurement was accurate to the nearest msec.

Procedure

Task orders

In all studies (A-D), the pair-recognition tasks were given in the first half of the testing session, followed by a 5-min. break, followed by the procedural learning task. Procedural learning was always given after pair-recognition tasks, so that temporal proximity effects (Chaiken, 1993) would work against the hypothesis that strength and resistance-to-interference abilities are consequential to late procedural learning performance. A temporal proximity effect is the tendency for two scores to be correlated with each other simply because the two scores have been observed close together in time. By giving procedural learning after the pair-recognition tasks, least practiced scores for procedural learning are closer in time to the pair-recognition scores than most practiced procedural learning scores. Thus, if there is a bias for pair-recognition to correlate to procedural learning owing to temporal proximity, this bias is stronger for the earliest procedural learning scores.

Pair recognition: Baseline and baseline+interference tasks

Each pair-recognition replication used different word materials, requiring new associations to be learned for each replication. Tasks were given in two replication sets with one baseline task and two baseline+interference tasks in each set. Twice as many baseline+interference tasks were given to equate diversity of materials over baseline and baseline+interference measurement (i.e. baseline+interference tasks use half as many words as baseline tasks). For half the participants, baseline+interference tasks were given before baseline tasks in each replication set, while for the other half, baseline tasks were given first.

Materials for tasks were selected without replacement from a pool of 48 3-letter nouns (i.e. ace, art, axe, box, bus, car, cop, cup, day, dog, dot, eel, eye, fog, fun, fur, hat, hip, hut, ice, ion, job, joy, log, lot, lox, nun, nut, oat, ore, owl, pen, pit, pot, sap, sin, sky, tea, ton, toy, vat, vip, war, web, wig, yak, yen, you). Words were randomly assigned to every condition and task replication for each participant. Each pair-recognition task used six pairs. Using unique letters to represent unique words, the following pair schemas were learned for a baseline task: AB, CD, EF (requiring right button responses) and GH, IJ, KL (requiring left button responses). For baseline+interference tasks the pair schemas were MN, OP, QR (right button) and MP, OR, QN (left button).

Participants initially studied the 3 right-button pairs in random order, twice, for 2.5 seconds a pair. Participants were told to memorize these pairs as "critical", to be later distinguished from "bogus" (hereafter, referred to as "foil") left-button pairs. Pairs were presented with one word (3.5 cm) below the other in a large lowercase font (i.e. 1 cm height), with the top word at the center screen. Specific words occurred at either the top or bottom position with equal frequency during study and test. Following study, participants were given a block of 24 pairs to classify in random order (4 replications of the 3 critical and 3 foil pairs). An error on a critical pair resulted in feedback and an additional 3 seconds of study for that pair. An error on a foil pair resulted in pair erasure and the message "Bogus Pair" for 1.5 seconds. The study/test procedure repeated two more times.

The participants were instructed to use the study/test blocks as their "grace period" in preparation for the real test blocks. These required 24 consecutive-correct responses to finish a set. If participants made an error before the end of the set, they restarted a new set of 24. Participants were required to do six such sets before leaving a pair-recognition task. Despite the fact that high accuracy was enforced, participants were also told that speed was important. In particular, at the end of each successfully completed set, the time to complete the 24 correct items was presented along with the participant's best score (either from the current or from a past set within the task). The participant's best score was identified as a "time to beat". Participants were then shown a histogram of their individual RTs, so they could see how their responses clumped together or spread apart along an ability scale where fast responses were high ability.

The average of median RTs for the 6 error-free sets is used as a score for each replication of a task. While this score has a disadvantage of reflecting different levels of practice for participants on these tasks, individual differences reflected by trials-to-criterion and error-rate scores from a specific task did not substantially overlap with the individual differences reflected by the RT score for that task. (This is an observation restricted to the current data and speed/accuracy sets and should not be construed as a general claim.) This was assessed by statistically removing-error rate and trials-to-criterion effects from the RT scores prior to their use in analysis and then comparing the results with the same analysis of uncorrected scores. Additionally, Pirolli and Anderson (1985) have shown equivalent performance when the amount of practice on specific materials is varied for a task similar to the current one (i.e. no difference between 12 and 24 repetitions per pair within a testing session). This suggests that the amount of practice at pair recognition can be varied between participants without necessarily varying the amount of memory strength reflected by the tasks.

Procedural learning task

The procedural learning task is from the “low-attention demand” condition of Woltz (1988), the rules for which were given before. Procedural learning uses numbers (between 1 and 20, excluding 10) as stimuli. Stimuli were presented in the same font as pair-recognition tasks. These stimuli were either uppercase words (e.g. "TWELVE", "FIVE") or digits (e.g. "12", "5"). Hence, word and digit stimuli differed in spatial extent as they did in Woltz (1988).

The basic teaching procedures of Woltz (1988) were used. These procedures included an instructional overview of the task, a 2-minute study period of the task rules, a demonstration of some representative problems, and explanatory error feedback throughout the task. Such feedback re-presented the rules relevant to the problem just received and derived the correct answer for the participant. Participants had to click a mouse button to leave the error feedback and resume the task.

Changes from Woltz (1988) were made to the end-of-block feedback in order to parallel the procedures described in the pair-recognition tasks, where a high-accuracy set was imposed. Participants were told they had a grace period of 7 blocks of 24 procedural learning problems before 24 consecutive-correct responses would be required to leave a set. At set 8 they were alerted that the grace period was over and that they needed to complete 25 more (error-free) sets. For the purposes of data analysis, consecutive stretches of 24 problems were used, not error-free sets. Hence, a given practice level on procedural learning does imply the same number of problems for every participant.

There were also minor cosmetic changes from Woltz (1988). The current version's font was larger. Stimuli were presented in upright (9 x 12 cm) rectangles centered in the screen. The properties UPPER half (2.5 cm from top) and LOWER half (2.5 cm from bottom) applied to the rectangles. Mouse rather than keyboard responses were given.

Stimulus balancing and randomization were adhered to within a completed error-free set. In particular, each set was balanced with respect to the number of different kinds of stimuli that could be presented (e.g. Digits, Big, Lower half). Stimulus randomization was in effect for all the partial sets. The specific stimuli were presented equally often across the task on average.

Results

Descriptive Data

Participants completed different numbers of procedural learning problems given variation in the number of set restarts each participant experienced. Therefore, an arbitrary block number had to be chosen, after which data would not be considered and before which a participant would not be considered. The block number, 40, was chosen because it maximized participant n and the amount of task practice all participants experienced. This amount of practice is 256 problems greater than for Woltz (1988). Using these criterion for exclusion, participant exclusion rate was relatively low (at least compared to Woltz, 1988, i.e. 13% vs. 22%).

Five procedural learning RT "epoch" scores were then computed by averaging 8 consecutive block medians for each. (Hence, the first RT epoch contains the grace period). However, for simplicity of presentation, only the first, third, and fifth epoch scores are used in structure modeling.

Basic mean data, broken down by pair-recognition task order (baseline task given before or after baseline+interference tasks), is shown in Table 1. Unlike procedural learning data, mean pair-recognition RTs are for the 6 error-free sets, whereas trials-to-criterion and percent correct apply to all the data. Significance levels are $p<.001$, unless otherwise noted. For the baseline+interference tasks, a within-subjects MANOVA found both a significant replication effect ($F(3,2427)=83$; $MSE=380486$), and a replication by order interaction ($F(3,2712)=8.7$; $MSE=143787$). The first baseline+interference task had longer RTs and that tendency was amplified when baseline+interference tasks were given before baseline tasks. For analyses comparing the two types of task to each other, RTs for critical (i.e. studied) and foil pairs were averaged over replications. A large task difference (i.e. interference effect) was found ($F(1,810)=8137$; $MSE=127916097$). Other significant effects were found (e.g. critical vs. foil differences and their interaction with type of task, baseline or baseline+interference); however, these do not bear on the issues discussed.

Table 1.

Descriptive data (means and standard deviations) for pair-recognition and procedural learning tasks by pair-recognition task order.

	Baseline First						Baseline+interference First					
TEST	RT		PC		TTC		RT		PC		TTC	
B1	684	109	96.3	3.6	288	80	680	87	98.2	1.3	264	49
B2	655	99	98.0	1.6	262	46	680	97	97.8	1.5	273	55
BI1	1104	220	92.0	5.2	362	114	1153	244	90.6	6.8	383	167
BI2	1056	196	95.6	2.4	311	71	1082	218	95.4	2.9	307	78
BI3	1034	187	95.4	2.4	314	75	1028	207	95.6	2.4	314	78
BI4	1061	206	95.5	2.6	314	74	1056	204	95.8	2.6	311	78
PL1	1603	434	88.9	9.4	.	.	1570	453	88.3	9.1	.	.
PL2	1325	327	95.3	2.6	.	.	1337	342	95.0	2.5	.	.
PL3	1197	262	96.3	1.9	.	.	1207	264	96.2	1.8	.	.
PL4	1123	220	96.5	2.0	.	.	1145	235	96.4	1.8	.	.
PL5	1088	206	96.6	1.9	.	.	1103	210	96.6	1.7	.	.

Notes. B1, B2 and BI1-BI4 are replications of the baseline and baseline+interference pair-recognition tasks, respectively. PL1-PL5 are learning epochs for a procedural learning task. Groups are separated by whether they received the baseline pair-recognition tasks first. RT = Reaction Time (msec); PC = Percent correct; TTC = Trials to Criterion. First and second columns of each score are means and standard deviations, respectively. Sample sizes: baseline first: n=402; baseline+interference first: n=409.

Correlational Results

Table 2 shows the partial correlations between baseline+interference replications and procedural learning performance epochs after controlling for the baseline task scores. This was computed in two ways. For the columns labeled "Raw", RT scores for pair recognition and procedural learning were used without taking into account participants' "error characteristics". These error characteristics are percent correct and trials to criterion for the pair-recognition tasks, and percent correct for the procedural learning epochs. For columns labeled "Adjusted", RT scores have been adjusted by partialing from each score all error characteristics from all scores. The results are highly similar despite the fact error characteristics, for a particular task, correlate significantly with the RT score for that task. The signs of the correlations (i.e. negative for accuracy and positive for trials-to-criterion) suggest stimulus-specific effects (i.e. some participants receiving harder pairs to learn than others in a pair-recognition replication).

The partial correlations in Table 2 provide empirical results relevant to the hypothesis that resisting interference in these tasks is both a unique ability and relevant to late skill. First, all observed partial correlations are significant which is consistent with uniqueness. Second, the dominant trend is that partials are larger with later procedural learning epochs. Recall that this relationship is expected given late-occurring procedural learning productions will have "fanned" components. However, there is another unanticipated trend, namely that the earlier baseline+interference tasks predict less well than the later ones.

Table 2.

Partial correlations for raw and accuracy-adjusted RTs: Baseline+interference (4 replications) against procedural learning (5 levels of practice) controlling for baseline task RTs.

	Raw				Adjusted			
	BI1	BI2	BI3	BI4	BI1	BI2	BI3	BI4
PL1	.23	.16	.19	.17	.21	.16	.19	.17
PL2	.24	.25	.28	.31	.24	.23	.30	.31
PL3	.25	.25	.33	.30	.26	.23	.34	.29
PL4	.21	.24	.34	.32	.23	.23	.35	.31
PL5	.22	.24	.35	.33	.22	.23	.36	.33

Notes. See Table 1 for column and row definitions. df on partials = 807, 789, for raw and adjusted respectively. For adjusted scores, each RT score was regressed on 18 scores and the residual score was used in the analysis. The 18 scores are accuracy and trials-to-criteria for pair-recognition tasks (12 scores), PL accuracy for each epoch (5 scores), and the between-subjects order variable. Rxx (split-half) for PL1 to PL5: .95, .97, .96, .96, .96, respectively. Rxx for BI1-BI4 and B1, B2 (alternate forms) may be derived for adjusted scores from Table A2.

Modeling Results

More details on Model 1

In Model 1, resistance-to-interference and the baseline contribute to the observed procedural learning between-epoch correlations. However, three pairwise-correlations among the epoch residuals (parameters represented by the bi-directional arrows joining pairs of es in Figure 1) also contribute. The epoch residual errors represent the part of procedural learning scores that is independent of the model (i.e. not explained by model factors). These are conceptually analogous to coefficients of alienation (Cohen and Cohen, 1975) in multiple regression.

The fact that there are correlations among the procedural learning e terms reflects the empirically large task-specific correlations among the procedural learning scores even after the effects of model factors are considered. Figure 1 indicates the large size of these correlations (minimum r=.45), and, in fact, their importance to overall model fit (e.g. the minimum $z > 10$ for these correlations) was too large to ignore in the modeling. In general, any large effect in the data that is left unrepresented in the model which is fit to the data, can result in a misleading set of "best-fitting" parameters. For instance, parameters representing a hypothesis (e.g. resistance-to-interference exists across baseline+interference and procedural learning tasks) may obtain values that refute the hypothesis, even while the hypothesis is true. This could happen because the parameters' best-fitting values account more for the (larger) unspecified effects in the model than their intended effects. Such, in fact, would happen in the current data. Therefore, correlating the errors of procedural learning scores is an important constant feature in every model considered in this paper.

Alternatively, one could choose to represent the task-specific variance in procedural learning as a procedural learning factor (either nested in the other factors or not). Models along these lines provide results that fully comport with results to be presented.

Model Adequacy

The 2-factor structure in Figure 1 (Model 1) was fit via the EQS program (Version 5.2, Bentler, 1993). Figure 1 shows the standardized measurement betas for adjusted score data. When the model is run on raw (unadjusted) data quantitative results are highly similar. (See Table 2 for definitions of adjusted and raw scores).

All paths from factors were "significant" (the least significant being resistance to interference on epoch 1 of procedural learning, $z = 4.8$). While significance levels depend on sample size, they are also model theoretic. A parameter is assessed as significant by how unlikely it would be for the estimate of that parameter to be zero (i.e. absent from the model) given the data under consideration and the maximum-likelihood estimation procedure (p-value determined by the normal z statistic, Bentler, 1993). The parameter for the variance of the resistance-to-interference factor was also significant (e.g. $z = 6.4$, for adjusted data; $z = 5.6$, for the raw data). If the baseline factor had been sufficient to explain the correlations among tasks, then the variance for this parameter (along with the paths from the interference factor to observed scores) should have been zero. However, more complete tests of the existence of an interference factor can only be provided by comparisons between specific models (next section).

The fit for Model 1 is also good. This can be seen by the high Bentler-Bonnet Nonnormed Index (BBNNI) given in Figure 1. Bentler (1993) suggests a minimum fit of .90 as cutoff level for adequate model description of the data. The BBNNI is a comparison between the "lack of fit" by the theoretical model and the lack of fit by a "null" model that posits no intercorrelations among scores (Hoyle and Panter, 1995, p. 166). Another fit statistic (provided along with the BBNNI) is the Root Mean Square Error of Approximation (RMSEA, Brown and Cudeck, 1993). The statistic's principal benefit is as a complimentary perspective on model fit, which is not based on the comparison to the "null" model. RMSEAs not exceeding .08 are desireable, with a fit value around .05 being ideal (a subjective opinion, Brown and Cudeck, 1993, p. 144). In any case, model fits, by themselves, are not as useful as comparisons between models. There may be some 1-factor models that provide "good" fit as I explore next.

Comparison of 1-factor and 2-factor models

The following analyses are for adjusted-score models, as parallel analyses for raw scores replicate the findings closely. Comparisons between 1 and 2-factor models are central to the current study. In particular, my goal is to show procedural learning and baseline+interference tasks are related to each other through a common interference factor that has independence from the baseline factor. It is critical that such interference be shown important for both baseline+interference and procedural learning tasks to demonstrate that the factor is not attributable to the unique (but reliable) variance specific to a single task. Hence, the one-factor models that I consider as strong challengers to the two-factor model are models in which one-factor explains the commonality between baseline+interference and procedural learning tasks, but other task factors may be entertained to explain correlations among replicates of the same task. Recall that in the two-factor approach (Model 1) there is already an implicit task factor for procedural learning (i.e. the correlations among the residual procedural learning errors). Therefore, the one-factor alternative models I assess remove interference from prediction of procedural learning but retain task-specific "interference" as an explanation for correlations among the baseline+interference replicates. There are two classes of models that accomplish this.

One class results from removing (i.e. fixing) free parameters from Model 1. This class of model is "nested" in Model 1 (i.e. Model 1 is a superset of this class of models) and can be compared to Model 1 using a Chi-square difference test. This test is a $\chi 2$ statistic derived from the difference in model chi-squares with df equal to the difference in model dfs. If the statistic is significant, so is the loss in model fit. For the nested class of alternatives, perhaps the strongest model to test would set the three paths from the interference factor to procedural learning scores to zero. The resulting loss of model fit from removing these paths was significant ($\Delta\chi 2(3) = 47$).

Another class of model is the set not strictly nested in Model 1. This set can be obtained by setting some Model 1 parameters to zero and also adding parameters that were not included in Model 1 (i.e. assumed zero in Model 1). Perhaps the strongest exemplar from this class is Model 1A shown in Figure 2.

MODEL 1A

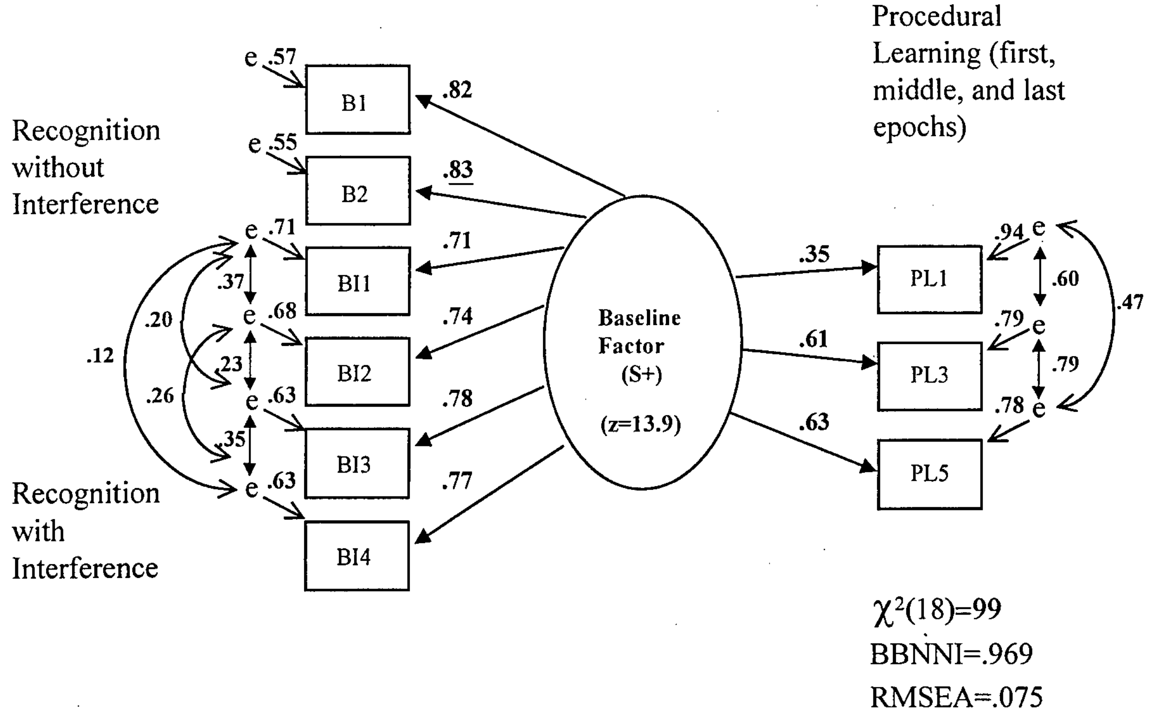

Figure 2.

A one-factor alternative to Model 1 which allows only the baseline factor explanation for pair-recognition and procedural learning correlations. All else is the same as in Figure 1. This model appears to fit as well as Model 1 because it accounts for an unanticipated effect in the data that is not represented in Model 1. When the two models are equated for this, Model 1 fits better (see text for discussion).

Model 1A also removes paths from the interference factor to procedural learning, but adds an implicit baseline+interference factor like the implicit procedural learning factor already present (in all models). This is accomplished by adding six pair-wise covariances among residual variances (i.e. correlating the e-terms) for baseline+interference tasks.

One can also compare non-nested models, provided the observed variables used in each model is the same (as in the current case). The approximate posterior probabilities for two competing models may be computed from differences in model chi-squares and number of free parameters used in each model (i.e. the complement of a model's degrees of freedom). This is the same information that the chi-square difference test uses; however, the result of the procedure is not a $\chi 2$ but a direct statement of likelihood of a proposed model in the context of other models and given the observed data. The full procedure and derivation is in Fornell and Rust (1989). That procedure is more general than its use here. For instance, there can be more than one alternative model and the effect of prior probabilities of the models (if known) can be weighted into the computation of likelihood. In the current assessment, I assume only one competing model (i.e. the non-nested Model 1A) and equal prior-probabilities for both models.

With this procedure Model 1's likelihood (relative to Model 1A) was estimated at only p=.035, which reflects the fact that Model 1A was observed to fit better ($\chi 2(18) = 99$) and have more degrees of freedom. However, the validity of this outcome can be questioned. Specifically, one general danger with non-nested model comparisons is that large extra-model effects can be left out of one model (i.e. Model 1) but included in the other (i.e. Model 1A). Therefore Model 1A may fit better (or as well) as Model 1 for a reason other than the sufficiency of a single factor expressing the commonality between pair recognition and procedural learning. Such an extraneous reason may be present in the non-nested part of Model 1A, i.e. the residual covariances among the baseline+interference tasks. While correlations among baseline+interference tasks are owing to the two model factors in Model 1, proximity effects may also be driving the observed correlations (c.f. Chaiken, 1993). Proximity effects could be expected for the first and the second pair of baseline+interference tasks, as these tasks are administered close together in time regardless of the order condition (baseline-first vs. baseline+interference-first).

When I investigated adding such proximity parameters to Model 1 by correlating the errors of only the adjacent baseline+interference tasks, the parameter for the first pair of baseline+interference tasks was highly significant ($\Delta\chi 2(1) = 57$). However, the parameter for the 2nd pair of baseline+interference tasks was not significant ($\Delta\chi 2(1) <1$). Adding just the first parameter causes the RMSEA fit index for Model 1 to drop from .079 to .049 (.090 to .056 in the raw data). Addition of this highly significant parameter also left the score loadings on resistance-to-interference (i.e. the factor betas for scores regressed on that factor) largely unchanged. The exception was the first two baseline+interference tasks loading less on the interference factor (especially the first baseline+interference score). When I compared Model 1 with the proximity parameter to Model 1A (which already has that parameter), the likelihood flipped in favor of Model 1 with likelihood not distinguishable from 1.0.

I also tested the idea that the proximity effect obscured model comparisons in another way. I fit both Models 1 and 1A (suitably modified) to a reduced set of scores that excluded the first baseline+interference score. This would remove the proximity effect from both Models 1 and 1A, while maintaining a fair comparison between models. When this was done, the goodness-of-fit comparison (as indexed by model chi-squares) clearly favored Model 1 (Model 1 $\chi 2(11)=15$, Model 1A $\chi 2(14)=74$), and again the posterior probability computed for Model 1 in the context of Model 1A was indistinguishable from 1.0.

Finally, there is yet another way to show, at least indirectly, that an interference factor is "needed" beyond Model 1A's "machinery". If one nests some extra machinery in Model 1A that reflects a general interference factor, then these extra parameters should not improve Model 1A's fit (given Model 1A's assumption of no general interference factor). Directly nesting all the interference parameters from Model 1 (i.e. the factor and its seven paths) will not work, because the model produced by this is not identifiable. However, another manifestation of a significant interference factor (between procedural learning and baseline+interference tasks) would be significant covariances between the e terms of baseline+interference and procedural learning scores, after Model 1A's parameters explained the data as best it could. Adding parameters for these (4 baseline+interference x 3 procedural learning) correlations results in an identifiable model, and that model fits significantly better than Model 1A ($\Delta\chi 2(12) = 81$, $p<.001$).

Resistance to interference: automatic or controlled processing ability?

The last two sections indicated that interference had some uniqueness from the baseline ability factor as well as some generality across pair recognition and procedural learning. Models that exclude an interference factor (by making it task-specific) fit the data more poorly than models with an interference factor that has some generality. Given this, one can consider interference ability in the context of skill acquisition. In particular, if procedural learning automates with practice, then interference ability could be considered to have automatic-processing characteristics if it has a stronger relation to later epochs of procedural learning than earlier ones. Conversely, a controlled-processing ability would be indicated, if that ability more strongly related to early procedural learning (c.f. Woltz, 1988). In fact, the model parameters (see Figure 1) indicate that resistance to interference has its greatest impact at later rather than earlier epochs; however, the difference between loadings for early and late epochs is not especially large (e.g. .24 vs. .32, respectively). Is this trend reliable or significant?

One approach to answering this question would be to bring a controlled-processing factor into the model to check whether the latent-structure analysis had sufficient power to detect a controlled-processing diminution with procedural learning practice. In this analysis a surrogate for controlled-processing ability is related to procedural learning, along with the baseline and interference factors. This model extended Model 1 to include a global controlled-processing factor in which all other factors were nested. In addition to assessing whether a controlled-processing decline can be detected, one can view the impact of having controlled-processing in the model on the interference factor's prediction of procedural learning. Under the assumption that interference and controlled processing overlap substantially, one would expect the predictive relation between interference and procedural learning to be decreased with controlled processing added to the model.

To put controlled processing in the model, four new tests from the Armed Services Vocational Aptitude Battery, or ASVAB, were added to the analysis. These tests were Arithmetic Reasoning, Math Knowledge, General Science Knowledge, and Word Knowledge, and were given some months before as part of the Air Force selection procedure. Kyllonen (1993) has shown that the general ability derived from the ASVAB is strongly correlated to a working memory factor derived from a battery of diverse information processing tests (i.e. $\underline{r}$ approaching 1). As working memory ability is often considered a strong correlate of controlled-processing ability (Ackerman, 1988; Woltz, 1988), the ASVAB tests provide plausible surrogates for controlled-processing ability.

The results for this model may be simply described. The predictive relationships between Model 1's factors and procedural learning were unchanged, both in magnitude and significance (e.g. procedural learning regressed on the interference factor were .24, .31, and .32, for 1st, 3rd, and 5th procedural learning epochs). In addition, the relation of the general factor to procedural learning declined as the skill was practiced as expected. For adjusted data, the weights were -.27, -.15, -.11 for 1st, 3rd, and 5th epochs. $\underline{Z}$ statistics were 6.5, 3.9, 2.9, respectively for the same. The weights are negative because a high achievement score goes with low reaction times. Both the different trends for the two factors (interference and controlled processing) and their independence of prediction indicate their distinctness as psychological factors. These results are inconsistent with a controlled-processing characterization of the interference ability introduced by the baseline+interference task. [2]

I have further investigated the plausibility of my controlled-processing "surrogate" because these markers are achievement tests and not information processing tests, per se. On different subjects than employed here, I have administered the baseline and baseline+interference tasks and derived baseline and resistance-to-interference factors on just the pair-recognition tasks. Parameter estimates and significance levels were similar to what is reported here for these tasks. Using task factors I have tried to predict the quantitative working-memory task from Kyllonen and Christal, 1990, which requires the subjects to perform simple arithmetic while maintaining a concurrent memory load. The working-memory task is arguably rich in controlled processing (c.f. Anderson, Reder, and Lebiere, 1996) but did not measurably load the interference factor (i.e. the regression beta for that factor was $r=.002$; $n=392$). However, consistent with my assumption that the general ability tests employed in the current study were, in fact, reasonable estimates of controlled-processing ability, the same working-memory task did load significantly on the general-ability factor defined from those tests ($r=.581$; $n=392$; $z=11.0$).

Discussion

Main results

Evidence for two distinct factors, a baseline and an interference factor, was found for practiced declarative and procedural recognition tasks. This evidence is embodied in the comparison of two-factor models to one-factor alternatives. In addition, both factors generally increased in importance with procedural learning with practice on the latter, while general ability showed a decreasing relationship.

Woltz (1988) found similar effects, namely a declining relation for working memory against procedural learning errors and an increasing relation for memory strengthening against procedural learning reaction time. However, in the current study, the different trends for controlled-processing ability and the baseline and interference abilities are observed within the same procedural learning performance scale (i.e. reaction time). Hence, one can conclude more strongly (than in Woltz, 1988) that the differing trends reflect different abilities. In the case where different trends are observed across different performance scales (i.e. RT and errors), such differences may also reflect the different characteristics of the measurement scales.

Effect size and robustness of the interference factor

The magnitude of the interference factor, or the percentage of variance it accounted for in observed variables (according to Model 1), was small compared to the baseline. (E.G. 9% of the variance in practiced procedural learning and 16% of the variance for later baseline+interference tasks. This is obtained by squaring the factor loading, or beta, which is similar to a semi-partial in multiple regression). The interference effect measured in reaction time is much larger than implied by the model apportionment of individual differences. That is the interference effect, as an RT difference, is about half the size of the baseline RT. However, the mean interference effect is not a pure index of resistance-to-interference ability. This can be shown empirically by correlating the interference effect (RT increment relative to the baseline) to the baseline RT ($r(809)=.36$, and $r(809)=.34$ for raw and adjusted data respectively).

It is a value judgement as to whether the interference loadings are large enough to be deemed "interesting" or of practical importance. However, these loadings are at least robust in the sense that they cannot be attributed to distributional abnormalities in the data or to the effect of outliers. I assessed this by re-running Model 1 on structures derived from scores under both a simple rank and a normalized-rank transformation of the individual data. Ranking completely changes the distributions of performance scores and also completely removes the distorting effects of outliers, if they exist. The models on the ranked scores replicate the findings that two factors are needed along with the general magnitude of the loadings. However, ranked models assigned the maximum effect of interference on the intermediate procedural learning epoch, and assigned larger effects on the first epoch (i.e. a relatively flat function rather than a monotonic increasing one).[3]

I've also looked at the variance/covariance matrix for reciprocal-transformed RTs (i.e. rates instead of RTs). This transformation does an excellent job at normalizing RT distributions, provided outliers are re-scored to the leading edge of the new (i.e. apparent) distribution. With or without (the eight) outliers, two-factor models would be found superior with similar parameters and significance levels. In all analyses (rates and ranks) the independence between interference and controlled processing was also observed.

Conceptualizing Resistance to Interference in Practiced Recognition

Relation to the Fan Effect

One could doubt that the current results bear on fan effects, because the baseline task seems qualitatively different from the propositional retrieval studied in fan-effect experiments, and the baseline+interference task at least looks similar to a "low-fan" condition in fan-effect experiments. However, the baseline task can plausibly be considered a zero-fan condition because every (experimental) path from probe words provides activation toward the correct response. When all paths send activation to the same response their effects can be assumed to summate (c.f. Jones and Anderson, 1987). With the interference manipulation added to the baseline, the task becomes a fanned condition, as it is implausible that foil pairs are not learned in this task. Therefore, each word in the baseline+interference probe has one irrelevant experimental association through which activation is lost. Presumably RT should increase for similar reasons as for the fan-effect paradigm, namely, activation loss down an irrelevant pathway. For this reason one cannot easily dismiss the current interference manipulation and the ability marked by it as being irrelevant to fan effects.

However, I speculate that the amount of activation in the relevant traces is not the only difference between baseline and baseline+interference tasks. An explanation of performance differences in terms of a unitary construct (e.g. amount of activation), would suggest that a one-factor model should have been sufficient. Conversely, it seems more reasonable, that given baseline+interference tasks can define a unique factor, that some qualitatively different types of memory processing are also involved in the interference manipulation.

Lack of strong relation to controlled processing

Given the interference manipulation is related to fan effects, the characterization of the factor as an automatic rather than controlled-processing ability is inconsistent with recent data on some fan effects. Conway and Engle (1994) found that "response competition" was sufficient for working memory (a.k.a. controlled processing) to correlate with fan effects (see also Cantor and Engle, 1993). In the current experiment, every baseline+interference probe word has the required response competition, so this task is in Conway and Engle's controlled-processing class of fan effect.

However, the discrepancy between the "automatic" character of the interference factor in this study and Conway and Engle's results may only reflect our differing methods. When I adopted their methods, I replicated their results. Specifically, when I compared participants from the 1st and 4th quartiles of a composite made from my general ability tests (as Conway and Engle did with their working-memory measure), participants low in general ability (or working memory) had a larger interference effect than people high in general ability. I also observed that this difference was reduced after general task practice. The difference between high and low ability interference effects was 74 msec initially (i.e. the average of the first pair of baseline+interference tasks minus the first baseline task) and was more than halved after general task practice (i.e. 32 msec for the average of the second pair of baseline+interference tasks minus the second baseline task; main effect of ability: $\underline{F}(1,402)=17.23$, $\underline{MSE}=574035$; interaction of ability with practice: $\underline{F}(1,402)=5.95$, $\underline{p}<.02$, $\underline{MSE}=88505$).

Hence, the baseline+interference task is sensitive to controlled processing as Conway and Engle found. However, the latent structure results do not indicate such sensitivity to be a significant part of the factor in common between baseline+interference and procedural learning tasks. The fact that both pair recognition and procedural learning are observed under a range of practice, probably allowed controlled-processing effects (early in tasks) to be separated from a more automatic factor in common with the practiced tasks.

Smaller interference effects (regardless of ability) were also observed for later replications of pair recognition (i.e. for the same extreme-groups analysis described above the effect of practice on "fan" effect was significant, $\underline{F}(1,402)=21.23$, $\underline{MSE}=315862$). Reduced fan effects for studied materials given initial practice on different materials has also been reported in Pirolli and Anderson (1985, Experiment 4). They interpreted this effect as a speed up in the "central processes" relevant to the fan effect (e.g. "comparison of the probe to memory" p. 151). However, controlled-processing resources, expended in becoming familiar with the fact-retrieval task (c.f. Ackerman, 1988) may also reduce the activation available for "spreading" thereby increasing the fan effect. However, such an interactive perspective on controlled processing and "automatic" activation processes (c.f. Anderson, Reder, and Lebiere, 1996) should not be confused with a perfect tradeoff between the two processing systems.

Resistance to interference in practiced recognition and "activation"

Resistance to interference, in the current study's context, arguably reflects a limitation of "activation" similar to that supposed for fan effects. However, Anderson, Reder, and Lebiere (1996) have also used the notion of activation limitations to model the effects of concurrent memory load on mathematical equation solving, a task arguably very rich in controlled processing. The characterization of both fan and working-memory effects as depending on activation limitations suggests a more apparent overlap between this study's interference factor and controlled processing ability.

However, despite the impression of a unitary activation in controlled-processing and practiced-recognition contexts, Anderson et. al.'s (1996) formulation of activation is not unitary. In particular, "source" activation refers to the resource limitation of working memory or "the salience or attention given to the [memory probe] cues" (p. 225). On the other hand, another sense of activation appears to cover processes linked to "controlling retrieval from declarative memory" (p. 225, see also p. 226 top). Anderson et. al. caution the reader that the two senses should be kept conceptually distinct. Both types of activation determine the total activation given a trace and therefore both types affect processing time. For lack of a better name from the ACT literature, I'll refer to the automatic type of activation as "historical", in the sense of depending on the frequency of the trace (i.e. memory strength) and the amount of overlap of the trace's components with other traces (i.e. fan). The automatic characterization of the interference ability found in this study and the lack of relation of interference ability to a working memory task indicate that the abilities underlying historical and source activation processes are distinct.

Skill specificity

The possibility that different types of recognition (procedural and declarative) depend on shared memory abilities, even after significant practice, is a counterexample to the skill-specificity hypothesis. This hypothesis purports that abilities underlying a task become less general with task practice. Given individual differences shrink with task practice (Hulin, Henry, & Noon, 1990; Ackerman, 1987; Fleishman and Hempel, 1954) or become less dominated by cognitive resources (Ackerman, 1988), skill-specificity is a natural conclusion. Because this issue is taken up in much greater detail in Experiment 2, I will postpone discussion of it here.

Experiment 2: The Unique Importance of the Memory-Strength Factor

Experiment 2 continues to explore the latent-structure methodology as a means of decomposing tasks into information-processing stages. As before, such decomposition is heuristic only, that is, it depends on the intuition that should a set of tasks exhibit a common ability with some uniqueness from other abilities this is fair evidence for a distinct "stage" in those tasks.

The main goal is to investigate the baseline factor, which is hypothesized to contain memory strength but other baseline abilities as well. In particular, the baseline factor is confounded with speed of motor responding (e.g. speed of selecting and executing the appropriate button click) and speed of letter/word operations. Motor ability is expected to increase in importance with skill practice (Ackerman, 1990), so the relative importance of memory strength is still moot.

The importance of a unique memory-strength factor, independent of these confounds, depends somewhat on one's perspective. Because memory strength is such a global and pervasive parameter in memory models, one is obligated to predict some importance for that concept as an individual difference (c.f. Underwood, 1975). In particular, memory strength appears to be a good candidate for an ability underlying automaticity. Anderson (1992, p. 170) explicitly states that the buildup of strength (for a production) is the "most important" construct with regard to ACT*'s explanation of automaticity. Similarly Logan (1990) draws an empirical and theoretical link between amount of repetition-priming (a memory-strengthening process) and the amount of automaticity exhibited in lexical decision performance. Recall also that Woltz (1988) has shown an empirical relation between repetition priming (his memory-strengthening measure) and late performance in the procedural learning task. Hence, memory-strength ability should be of consequence to skill learning, and of particular importance late in learning where automaticity has developed. Finding uniqueness for memory strength from the other baseline processes would therefore support cognitive theories of skill acquisition.

However, from another perspective automated performance might be cognitively lean. At least for some theories (Ackerman, 1988, 1990), the automatic phase is associated with motor abilities and not cognitive ones. This perspective tends to view individual differences in cognitive abilities as negligible in the automatic phase. Hence, memory strength deconfounded from motor ability should show relatively little importance to late procedural learning (at least when compared to motor ability).

Model 2 (Figure 3) is a latent-structure model that is relevant to the above perspectives. This model extends Model 1 by decomposing the baseline factor of Model 1 into memory-strength, letter-word processing, and motor factors. Notice that the memory-strength factor spans (i.e. has arrows to) only the learning tasks, while the letter-word processing factor spans two replications of a lexical decision task and the learning tasks. These model specifications embody a hypothesis that lexical decision indexes a participant's ability for processing double-word displays (e.g. reading speed, semantic-memory retrieval speed), but that such ability does not depend on the memory-strength and resistance-to-interference abilities in the learning tasks. Also inherent in these specifications is the hypothesis that the learning tasks will depend on the letter-word factor in addition to memory strength.

These specifications implement a similar analysis to Experiment 1's. That is, the model first estimates the reliable letter-word processing variance within lexical decision and learning tasks (i.e. by defining a factor for all tasks that involve letter-word processing). Then another part of the model assesses the common variance left over among learning tasks, after accounting for letter-word processing ability. If the memory-strength factor is still needed to explain the residual correlations among learning tasks, then the learning tasks will still load significantly on that factor.

In a similar fashion, the choice-reaction time task can be used as an index of a motor-processing and response-selection factor common to all tasks. All previous factors can be nested within this new baseline factor and similar hypotheses assessed. That is, one can assess whether the letter-word processing and the doubly nested memory-strength factors are still needed after participants' motor abilities have been accounted for.

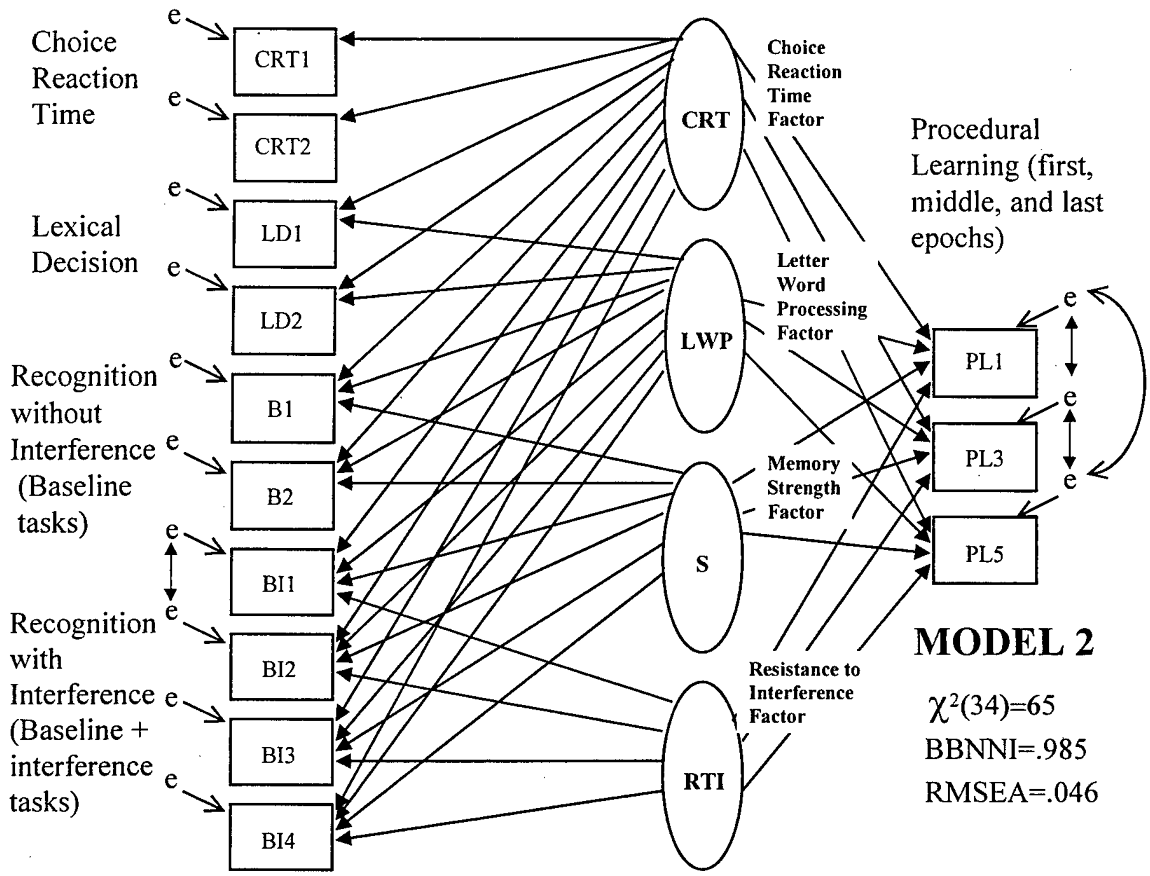

Figure 3.

An expanded latent-structure model for practiced recognition that separates memory strength from letter/word processing and motor abilities, and also includes a factor for resisting interference. Fit statistics are as in Figure 1. Path coefficients and other model results are presented in Table 3.

Method

Participants

Participants were 505 Air Force recruits (92% male; 8% female) taken from the sample of "Experiment 1". About 12% were not analyzed for reasons discussed before. Hence 434 participants were analyzed.

Procedure

General

This experiment's data is a subset of Experiment 1's, so the same equipment and procedures for pair recognition and procedural learning were used. Therefore, I only describe the tasks unique to Experiment 2.

Choice-reaction time task

Three pound-signs, "###", were presented either on the left or right side of the screen. Right and left stimuli required right and left mouse clicks, respectively. The stimulus was 1.5 cm in height and 2 cm in length and displayed 8.5 cm either left or right of the column center in the center row of the screen. Feedback and accuracy sets were the same as in Experiment 1 tasks.

This task was given in two replications. Replication 1 had a grace period of three practice blocks of 24 trials (balanced with respect to stimuli), after which 6 error-free sets were required of the participant. Replication 2 had no grace period and also required 6 error-free sets. Data used in modeling are only from the error-free sets. Two markers of the motor and response-selection factor were constructed by averaging time 1 right-response RT and time 2 left-response RT for marker 1, and the complementary set for marker 2.

Lexical decision task

Lexical decision trials employed the same display format as the pair-recognition tasks. For each participant, fifteen words were randomly selected from the population of words used in pair-recognition tasks with replacement (meaning some words appearing in the lexical decision task also appeared in later pair-recognition tasks). The 15 words defined a set of 105 distinct word-word pairs (hereafter WW pairs). Out of the WW pairs, 105 nonword-word pairs (hereafter NW pairs) were created by first permuting the positions of the WW words and then randomly selecting one word from the pair to be transformed to a nonword by vowel substitution.

Participants did 8, stimulus-balanced, blocks of 26 lexical decision problems (all but two of the constructed problems). The entire test was a different random sequence for each participant with the constraint that related WW-NW problems occur in different test halves. For instance, if wig/tea were given to the participant in the first half, later either toa/wig or tea/wug would be given in the second half. If either of the two NW pairs were given in the first half, then wig/tea was given later. WW trials required a right-click response and NW trials required a left-click.

Because lexical decision problems never repeated, the methodology, which enforces accuracy by requiring error-free sets, could not be used. Instead accuracy was stressed as being important and accuracy-corrected RT feedback was given for each block. The accuracy-corrected feedback expresses participant RT performance as the sum of the response times divided by the number correct after correcting for guessing. Participants were encouraged to answer quickly by identifying their best accuracy-corrected RT score from the current or previous sets as the time to beat.

Each block's WW and NW median RT was computed. Averaging NW conditions from odd blocks with WW conditions from even blocks and averaging the complementary set of conditions formed the two lexical decision markers.

Task Orders

Choice-reaction time replication 1 and the lexical decision task were the first and second tests of the session, respectively. The first replication set of pair-recognition tasks followed (i.e. one baseline and two baseline+interference tasks). Next the second replication of choice-reaction time and the second replication set of pair-recognition tasks followed, respectively. Finally, there was a 5-min break and the procedural learning task.

Results

Descriptive and Correlational Data

I only briefly describe the new tasks. The choice-reaction time task showed fast RT and high accuracy (291 msec and 98.6%, average median RT and accuracy over all conditions). The lexical decision task was considerably slower and less accurate (1044 msec and 91% correct, average median RT and accuracy over all conditions). While accuracy correlates to speed in both tasks, they do not correlate the same way ($r(432)= .23$ and $r(432)= -.33$, $ps<.001$, for choice-reaction and lexical decision, respectively). As before, I statistically removed speed/accuracy effects from the RT scores and compared results for analyses on adjusted and raw RT scores.

Modeling Procedure and Results

Unfortunately the proximity effect that obscured model comparisons in Experiment 1 also affected Experiment 2 results. I will outline my analysis procedure on the adjusted data, as the analyses with the raw data were closely parallel. As the presence or absence of the proximity effect in the model only qualitatively affected the presence of the interference factor, I will only provide descriptive information relevant to the interference factor.

The first structure model fit was a version of Model 2, which did not account for the proximity effect (i.e. removed the correlated errors between the first two baseline+interference tasks that is shown in Figure 3). In this model all paths from factors to observed variables were free parameters and all factor variances (e.g. choice-reaction time, strength) were fixed at 1.0. The fit for the model was $\chi2(35)=95$ with interference betas for procedural learning epochs (1, 3, 5) estimated at .10, .01, -.06, respectively, and for baseline+intereference tasks (1-4) estimated at .67*, .19*, -.01, -.12, respectively. Notice the two high loadings (where the asterisks indicate a $z>3.0$) reflect the proximity effect on the first two baseline+interference tasks. To test whether this effect had usurped the intended function for the interference factor, I next fit Model 2 adding a correlation between the errors of the first two baseline+interference tasks. As this model takes into account the proximity effect, the interference factor is "free to come back" to improve model fit, if it's needed. This model's fit was $\chi2(34)=65$ (a significantly better fit via a chi-square difference test). The interference betas provided by this model were .06, .23*, .28*, for procedural learning epochs (1, 3, 5), respectively, and .18, .32*, .43*, .48* for the baseline+intereference tasks (1-4), respectively, which is consistent with Experiment 1 results. Finally, I fit a version of Model 2 that removed the proximity effect by excluding the first baseline+interference score from the analysis. This model's fit ($\chi 2(27)=40$) is not directly comparable to the other model fits (owing to the different number of observed scores); however, the interference betas were .08, .25*, .29* for procedural learning epochs (1, 3, 5) and were .32*, .44*, .48* for baseline+interference tasks (2-4). The fact that the betas are very similar for models that account for or exclude the proximity effect indicates that the measurement of the interference factor was obscured when the proximity effect was not accounted for.

The final latent structure model reported is the model with the proximity parameter included (i.e. Model 2 in Figure 3). The standardized solution (reflecting the best fitting parameters for the latent structure model) is given in Table 3. This model also fixed one factor path to 1.0 for each factor, allowing each factor's variance to be estimated as a free parameter. A reasonable rule of thumb is to fix the path for each factor that has the highest standardized loading on that factor (as determined by the model in which all paths are free and factor variances are all set to 1.0). Fixing a path (and freeing the variance) or fixing the factor (and freeing all paths) are mathematically equivalent models, although the initial starting values for model parameters may have to be different in order for the two models to converge to the same solution.[4]

Table 3.

Standardized measurement equations for Model 2.

			Adjusted		
Score	CRT	LWP	S	RTI	e
PL1	.20	.21	.14[a]	.06[n]	.95
PL3	**.34**	**.29**	.29	.23	.81
PL5	**.35**	**.32**	**.36**	.28	.75
B1	**.47**	**.39**	**.60**		.52
B2	**.46**	**.33**	.63[F]		.53
BI1	**.31**	**.37**	**.48**	.18[a]	.71
BI2	**.34**	**.38**	**.50**	.32	.62
BI3	**.31**	**.36**	**.55**	**.43**	.53
BI4	.28	**.39**	**.55**	.49[F]	.47
LD1	**.40**	**.87**			.30
LD2	**.38**	.91[F]			.16
CRT1	**.95**				.32
CRT2	.98[F]				.19

Notes.
See Figure 3 for factor/score definitions and model fit statistics. Bold fonts and regular fonts are significant $z>6.0$ and $z>3.0$, respectively. F, a, and n superscripts indicate a fixed parameter (path), significant at $z>2$, and not significant, respectively. Model zs for RT, LWP, S, and RTI variances (respectively): 13.1, 13.0, 8.0, 4.9. Residual correlations r(PL1,PL3), r(PL1,PL5), and r(PL3,PL5) are estimated at .59, .44, .78 (zs>6). The proximity effect, or the residual correlation r(BI1, BI2), is estimated at $r=.40$ ($z>6$). $n=434$.

Finally, when Model 1, with a proximity parameter, is fit to Experiment 2's data, loadings on the interference factor are very close to the ones in Table 3 (i.e. .06, .22, and .27 for procedural learning and .19, .33, .45, and .51 for baseline+interference tasks). This result provides an interesting demonstration, namely that expanding nested factor models to include more factors (and approximately doubling the number of free parameters) does not necessarily obviate the need for, or change the qualitative patterns of, factors which are nested in the new model factors.

Discussion

Main Results

Experiment 2 reflects a better design by including measurement of other hypothetical information-processing stages that could reside in the baseline factor of Experiment 1. When memory strength was deconfounded from these other abilities, the factor was still evident. Additionally, memory strength increased in importance against procedural learning with practice on the latter as expected by skill-acquisition theories where memory strength is an important rate limiting factor (e.g. Anderson, 1987; Anderson, 1992; Logan, 1990). Resistance to interference had similar effects in the more inclusive information-processing model of procedural learning.

Other ability factors

The choice-reaction time task marked a significant factor in common with practiced procedural learning (and pair recognition). This is expected from skill-acquisition theories that implicate motor abilities in later stages (e.g. Ackerman, 1988, 1990). The interesting news in the current study, however, is that rather than dominating prediction, motor ability's contribution was similar to the three other cognitive factors identified. Of course, one can always suggest that with more practice motor abilities would eventually dominate.

The lexical decision task also marked a significant factor in common with practiced procedural learning (and pair recognition). This suggests either access to semantic memory (as might occur in parity or magnitude judgments in procedural learning) or letter/word processing stays important in procedural learning performance as the skill tends towards automaticity.

More on Skill Specificity

An important trend in Table 3 (which also occurs in Figure 1) is the decreasing error residual for the later procedural learning epochs. In other words, the later epochs are being predicted better than earlier ones by the ability factors I investigated. The internal reliability of the procedural learning epochs cannot account for this effect because they are static (see Table 2 notes). This obviously contradicts the skill-specificity hypothesis introduced before which states that general individual differences decrease with task practice and are replaced by task-specific variance.

Ackerman (1990) thought it important to demonstrate counterexamples to the skill-specificity hypothesis. In particular, he showed a practiced motor/aiming task increased in communality to simulated air-traffic controller performance, as the latter became increasingly practiced and more dependent on operating the controls than strategic thinking. Hence, motor/perceptual abilities are not encapsulated to the task they are learned but reflect individual differences general to different tasks. The current results extend Ackerman's counterexample by showing the skill-specificity hypothesis can be contradicted in a largely cognitive domain by memory-related abilities (e.g. stable letter-word processing proficiencies, memory strength, and interference).

However, the idea of a factor's increasing "importance" to a task with task practice should be carefully considered. The importance of a factor when conveyed by the standardized measurement equations (which is what I report and what is typically reported) indicates how much the factor influences a score relative to that score's standard deviation. However, the importance as conveyed by the unstandardized equations indicates how much of the factor is used in an absolute sense (i.e. the unstandardized equations express observed test scores in terms of linear functions of the factors). In contrast to the standardized factor betas, the unstandardized B weights can stay the same or even decrease for later procedural learning epochs. Therefore, when a factor increases its "importance" to a task with task practice this can also mean that the factor is static in its importance throughout learning. However, such factors may account for proportionally more task variance, later in practice, because initial performance drivers (e.g. controlled processing) have become less important.

General Discussion: What a Latent-Structure Perspective Contributes to Cognitive Psychology

The primary methodology that allowed separation of interference and memory strength in recognition comes from a specific type of model used in confirmatory factor analyses, called a "nested factor" model (Gustafsson and Balke, 1993). As Experiment 2 shows, a strict nesting of experimental conditions (as occurred only for pair-recognition tasks) is not a requirement; though strict nesting makes the psychological interpretation of the factor more defensible.

Nested-factor modeling is a good alternative to exploratory factor analysis applied to a diverse set of memory tasks. The latter method has not proven very powerful at isolating different memory abilities or systems (e.g. Malmi, Underwood, and Carroll, 1979; Underwood, Boruch, Malmi, 1978). Nested-factor modeling might also be viewed as a practical complement to estimating (different) abilities via cognitive models of performance. This approach uses parameter values from a theoretical performance function fit to each subject as the individual-differences measures (Lohman, 1994; Jensen, 1987; Sternberg, 1977).

One interesting parallel between psychometric methods, as embodied by nested-factor modeling, and cognitive psychology is the emphasis on stage decomposition. It is my belief that nested-factors models allow a proof-by-construction method for determining whether a task operation reflects a "stage". If an experimental manipulation (or a set of tasks) introduces a new ability with measurable independence from other abilities, this would seem at least as diagnostic and intuitive as other methods proposed for stage identification (e.g. additive factors, Sternberg, 1969). However, the failure to find any uniqueness between two proposed stages, in the individual-differences sense, has no bearing on whether the stages are distinct. It is logically possible for two distinct stages to depend on the same sorts of processing ability. The individual-differences heuristic for a stage, that I consider here, also makes no claims about serial/parallel or stage ordering issues.

The lesson's learned with a latent-structure perspective can comport with the assumptions of cognitive psychology or be surprising. The finding of distinct strength and interference factors for procedural and declarative recognition tasks support a general assumption that strength and interference are distinct limitations of such tasks. This finding was expected given nomothetic perspectives of such tasks. Another finding supportive of current cognitive perspectives was the demonstration that memory-strength ability is at least as important to late skill performance as other abilities (c.f. Woltz, 1988).

However, the lack of correlation between activation processes in practiced recognition and controlled-processing ability could be construed as surprising, at least relative to specific literatures (e.g. Cantor and Engle, 1993). A strong belief for the overlap between these two processing domains might have been expected given the central place for activation in some unified cognition theories (e.g. ACT*, Anderson, 1983, although see Anderson, Reder, Lebiere, 1996 which is more ambiguous on this issue). However, the activation ability, defined and investigated here, would apparently not extend to working memory tasks. Therefore, the common use of term "activation" in working memory contexts (e.g. Anderson and Matessa, 1997) and in practiced skill contexts is misleading. In summary, despite the fact the effect of "activation limitation" can be modeled in highly similar ways (i.e. behave homologously) across the two contexts, the fact that they can is not a test of their sameness, as the individual-differences data can clearly show.

References

Ackerman, P. L. (1987). Individual differences in skill learning: an integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Ackerman, P. L. (1990) A correlational analysis of skill specificity: Learning, abilities, and individual differences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 883-901.

Anderson, J. R. (1981). Interference: The relationship between response latency and response accuracy. Journal of Experimental Psychology: Human Learning and Memory, 7, 326-343.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill Acquisition: Compilation of weak-method problem solutions. Psychological Review, 94, 192-210.

Anderson J. R. (1992). Automaticity and the ACT* theory. American Journal of Psychology, 105, 165-180.

Anderson J. R. (1993). Rules of the mind. Hillsdale, NJ, Lawrence Erlbaum Associates.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working Memory: Activation Limitations on Retrieval. Cognitive Psychology, 30, 221-256.

Anderson, J. R & Matessa, M. (1997). A production system theory of serial memory. Psychological Review, 104, 728-748.

Bentler, P. M. (1993). EQS: structural equations program manual. Los Angeles, CA: BMDP Statistical Software.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds), Testing structural equation models (pp. 136-162). Thousand Oaks, CA: Sage Publications Inc.

Cantor, J., & Engle, R. W. (1993). Working memory capacity as long-term memory activation: an individual-differences approach. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1101-1114.

Chaiken, S. R. (1993). Test-proximity effects in a single-session individual-differences study of learning ability: The case of activation savings. Intelligence, 17, 173-190.

Cohen, J. & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Conway, R. A. & and Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. Journal of Experimental Psychology: General, 123, 354-373.

Cudeck, R. (1989) Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.

Donaldson, G. D. (1983). Confirmatory factor analysis models of information processing stages: An alternative to difference scores. Psychological Bulletin, 94, 143-151.

Dosher B. A. (1981). The effects of delay and interference on retrieval dynamics: Implications for retrieval models. Cognitive Psychology, 13, 551-582.

Dosher B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 541-574.

Fleishman, E. A. & Hempel, W. E. (1954). Changes in the factor structure of a complex psychomotor test. Psychometrika, 19, 239-252.

Fornell, C. & Rust, R. T. (1989). Incorporating prior theory in covariance structure analysis: a Bayesian approach. Psychometrika, 54 (2), 249-259.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Gustafsson, J. E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. Multivariate Behavioral Research, 28, 407-434.

Howe M. L. (1995). Interference effects in young children's long-term retention. Developmental Psychology, 31, 579-596.

Hoyle R. H., & Panter A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 158-176). Thousand Oaks, CA: Sage Publications Inc.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships, Psychological Bulletin, 107, 328-340.

Jensen, A. R. (1987). Process differences and individual differences in some cognitive tasks. Intelligence, 11, 107-136.

Jones, W. P. & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. Journal of Experimental Psychology: General, 116, 137-153.

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-source model. Journal of General Psychology, 120, 375-405.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! Intelligence, 14, 389-434.

Kyllonen, P. C. & Tirre, W. C. (1991). Knowledge and processing speed determinants of associative learning. Journal of Experimental Psychology: General, 120, 57-79.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? Cognitive Psychology, 22, 1-35.

Long, S. J. (1983). Covariance structure models: an introduction to LISREL. Beverly Hills, CA: Sage Publications.

Lohman, D. F. (1994). Component scores as residual variation (or why the intercept correlates best). Intelligence, 19, 1-11.

Malmi, R. A., Underwood, B. J., & Carroll, J. B. (1979). The interrelationships among some associative learning tasks. Bulletin of the Psychonomic Society, 13, 121-123.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. (pp. 1-55), Hillsdale, NJ: Erlbaum.

Pirolli P. L. & Anderson, J. R. (1985). The role of practice in fact retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 136-153.

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum.

Sternberg, R. J., & Gastel, J. (1989). Coping with novelty in human intelligence: an empirical investigation. Intelligence, 13, 187-197.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. In W. G. Koster (Ed), Attention and Performance II. (pp. 276-315), Amsterdam: North-Holland.

Tirre, W. C. & Pena, C. M. (1993). Components of quantitative reasoning: General and group ability factors. Intelligence, 17, 501-521.

Underwood, B.J. (1975). Individual differences as a crucible in theory construction. American Psychologist, 30, 128-134.

Underwood, B.J., Boruch, R.F., & Malmi, R.A. (1978). The composition of episodic memory. Journal of Experimental Psychology: General, 107, 393-419.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. Journal of Experimental Psychology: General, 117, 319-331.

Woltz, D. J. & Shute, V. J. (1993). Individual differences in repetition priming and its relationship to declarative knowledge acquisition. Intelligence, 17, 333-359.

Yee, P., Hunt, E., & Pellegrino, J. W. (1989). Coordinating cognitive information: Task effects and individual differences in integrating information from several sources. Cognitive Psychology, 23, 615-680.

Acknowledgments

This research was conducted by personnel of the Learning Abilities Measurement Project (LAMP), which was funded by the United States Air Force (AFOSR and Armstrong Laboratory; WU: 2313HM15, 2313T154). Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Air Force. I am indebted to Phil Ackerman and many anonymous reviewers, whose comments helped to define the paper's focus and greatly improve the presentation. I'd also like to thank Ginger Goff, Pat Kyllonen, Bill Tirre, for theoretical discussions and Rich Walker and Janice Hereford for technical support and data management. Scott Chaiken is now at the Air Force Research Laboratory, AFRL/HEAI, Brooks AFB, TX, 78235-5118 (email: scott.chaiken@brooks.af.mil).

Footnotes

1. A "pure interference" task that does not depend on memory strength would seem unlikely, as strength is ubiquitous in the modeling of memory retrieval (e.g. SAM and ACT-type theories). However, a pure "memory strength" task might also be hard to defend. Graphemic, orthographic, and semantic similarity of words might provide uncontrolled sources of interference. Also pre-experimental word associations and within and between-task effects (e.g. list-length, and proactive inhibition) might also contribute. While I can't dismiss the possibility of such effects, I can note that they are equally represented in the baseline and baseline+interference tasks. Hence, if such effects were large and of similar nature to the interference ability defined by Model 1, there should be no variance left over in an observed score after prediction by the baseline factor. One might also expect some weakening of an interference factor, defined by the experimental manipulation, in proportion to the amount of uncontrolled interference effects in the baseline.
2. This analysis did not include or depend on the proximity parameter between the first two baseline+interference replications, and results replicate using scores unadjusted for accuracy and trials-to-criterion data. For adjusted scores, $\chi 2(45)=164$; $p<.001$, Bentler-Bonett Nonnormed fit = .966. Root Mean Square Error of Approximation = .057. Model zs for S^+, RTI, and general ability variances: 14.3, 6.4, 10.2, respectively. Finally, for the general ability tests, the general science and word knowledge tests were allowed to have correlated errors to account for a large residual correlation between them (i.e. a verbal ability factor nested in general ability).
3. Ranking scores removes the information present in the variance of the scores (i.e. the resultant standard deviation for every score becomes a function of sample size). For many models it is inappropriate to model score correlations with equal standard deviations for every score. However, with "fully nested" models estimated via maximum likelihood (the approach of Model 1 and 2), results on correlation matrices of z-scored variables (i.e. scores with equal standard deviations) will be the same as the results on full covariance structures (Krane and McDonald, 1978 cited in Cudeck, 1989).
4. Here I report Study D (n=478) which differs from Experiment 2 by: 1) using Pentium machines and 2) using a lexical decision task that employed all 48 3-letter nouns in 288 problems (as opposed to just a subset of 15 nouns in 208 problems). I report Model 2 fit on adjusted data (as raw data results qualitatively replicate). Betas for the 4 factors for procedural learning (epochs 1, 3, and 5, respectively) were: CRT(.19, .30, .32), LWP(.32, .37, .37), S(.21, .37, .42), RTI(.22, .29, .28), for baseline+interference tasks (BI1-BI4, respectively): RTI(.31, .32, .42, .28). The above betas were significant with $z>3.0$. Factor variances had zs of 15.4, 13.9, 8.8, 4.1, for CRT, LWP, S, and RTI, respectively. The proximity parameter was significant at $z=2.7$ ($r=.16$). Fit was $\chi 2(34)=62$; $p<.003$.

Appendix A: Correlation Matrices and Standard Deviations

Table A1.

Correlations and standard deviations relevant to Experiment 1 for raw scores (top) and scores with effects of error rate and trials-to-criterion removed (i.e. adjusted scores, bottom).

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.	PL1	----	.64	.56	.30	.25	.36	.30	.32	.31	-.27	-.18	-.17	-.12
2.	PL3	.66	----	.87	.46	.44	.47	.48	.54	.52	-.19	-.22	-.12	-.08
3.	PL5	.56	.87	----	.48	.48	.47	.49	.57	.56	-.14	-.17	-.12	-.06
4.	B1	.26	.47	.48	----	.70	.60	.57	.58	.58	-.07	-.16	-.12	-.12
5.	B2	.24	.45	.48	.72	----	.55	.61	.62	.64	-.06	-.16	-.10	-.10
6.	BI1	.33	.49	.47	.60	.56	----	.68	.58	.55	-.11	-.16	-.14	-.16
7.	BI2	.29	.48	.49	.60	.61	.70	----	.63	.64	-.09	-.20	-.11	-.14
8.	BI3	.32	.55	.58	.61	.64	.64	.67	----	.71	-.09	-.18	-.08	-.07
9.	BI4	.30	.52	.56	.60	.65	.60	.68	.74	----	-.08	-.15	-.06	-.09
10.	ARITHRES	-.24	-.13	-.09	-.02	-.04	-.02	-.03	-.04	-.04	----	.60	.37	.30
11.	MATHKNOW	-.15	-.13	-.10	-.11	-.12	-.09	-.15	-.12	-.10	.60	----	.32	.21
12.	GENSCI	-.14	-.06	-.07	-.05	-.06	-.05	-.06	-.04	-.03	.37	.32	----	.54
13.	WORDKNOW	-.09	-.01	-.00	-.07	-.06	-.08	-.08	-.03	-.04	.30	.21	.54	----

Table A1. (continues).

Standard Deviations:

Raw: 444.0 262.8 207.7 98.5 98.5 233.8 207.6 197.3 205.1 6.4 6.4 6.4 4.2.

Adjusted: 428.8 244.0 200.0 88.3 92.6 193.6 183.2 180.2 191.0 6.4 6.4 6.4 4.2.

N=811 for scores 1 - 9. N=809 for scores 10 - 13. See Figure 1 and text for score description.

Table A2.

Correlations and standard deviations relevant to Experiment 2 for raw scores (top) and scores with effects of error rate and trials-to-criterion removed (i.e. adjusted scores, bottom).

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.	PL1	----	.63	.53	.33	.26	.35	.31	.30	.28	.33	.30	.22	.25
2.	PL3	.64	----	.88	.47	.42	.45	.47	.51	.51	.41	.41	.34	.36
3.	PL5	.52	.86	----	.52	.49	.45	.50	.58	.57	.45	.44	.36	.37
4.	B1	.26	.46	.51	----	.71	.60	.60	.56	.60	.55	.55	.48	.51
5.	B2	.23	.41	.48	.72	----	.53	.61	.58	.63	.49	.51	.46	.48
6.	BI1	.29	.43	.44	.59	.54	----	.71	.55	.52	.50	.50	.29	.32
7.	BI2	.27	.44	.48	.60	.60	.72	----	.61	.65	.50	.50	.35	.38
8.	BI3	.26	.48	.57	.61	.60	.61	.65	----	.71	.40	.43	.33	.32
9.	BI4	.22	.48	.55	.60	.63	.56	.69	.74	----	.45	.47	.31	.31
10.	LD1	.28	.40	.43	.52	.46	.45	.46	.42	.44	----	.95	.40	.43
11.	LD2	.26	.39	.42	.53	.48	.45	.48	.45	.47	.94	----	.38	.41
12.	CRT1	.18	.32	.34	.45	.44	.29	.32	.31	.29	.38	.35	----	.94
13.	CRT2	.20	.33	.34	.46	.45	.30	.34	.30	.27	.39	.37	.93	----

Table A2. (continues).

Standard deviations:													
raw:	462.9	273.2	207.1	103.5	101.8	257.0	217.9	201.8	205.7	168.2	165.1	29.8	29.2
adjusted:	432.1	238.8	190.8	88.5	94.4	207.5	187.9	182.2	186.8	146.0	144.3	28.0	27.4

N=434. See Figures 1 and 3 for score description.